Enhancing Revisitation in Touchscreen Reading for Visually Impaired People with Semantic Navigation Design

ZHICHUN LI, Department of Computer Science and Technology, Tsinghua University, China
JIANG YU, Department of Computer Science and Technology, Tsinghua University, China
XIAOCHEN LIU, Department of Cognitive Science, University of California, San Diego, United State
YUHANG ZHAO, Department of Computer Science, University of Wisconsin-Madison, United State
CHUN YU*, Department of Computer Science and Technology, Tsinghua University, China
YUANCHUN SHI, Department of Computer Science and Technology, Tsinghua University, China

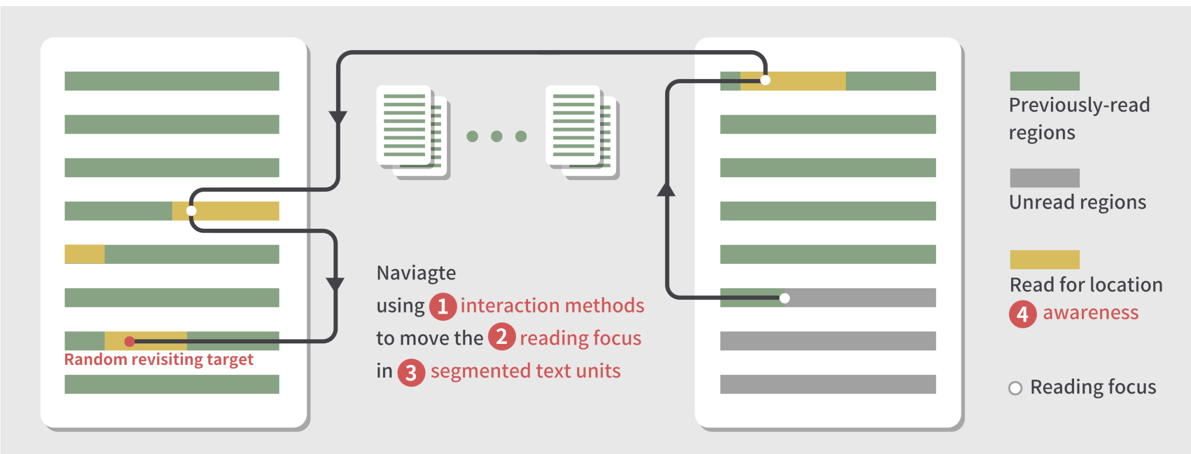

Fig. 1. We summarized four factors of the visually impaired readers' revisiting process on touchscreens through an observation. To facilitate VI revisiting, we proposed design guidelines by optimizing the four factors and implemented a prototype SRVI.

Revisitation, the process of non-linearly returning to previously visited regions, is an important task in academic reading. However, listening to content on mobile phones via a screen reader fails to support eyes-free revisiting due to its linear audio stream, ineffective text organization, and inaccessible interaction. To enhance the efficiency and experience of eyes-free revisiting, we identified visually impaired people's behaviors and difficulties during text revisiting through a survey (N=37)

*This is the corresponding author.

Authors' addresses: Zhichun Li, thulizc@163.com, Department of Computer Science and Technology, Tsinghua University, Beijing, China; Jiang Yu, Department of Computer Science and Technology, Tsinghua University, Beijing, China, jiangyu0927@gmail.com; Xiaochen Liu, Department of Cognitive Science, University of California, San Diego, San Diego, California, United State, xil001@ucsd.edu; Yuhang Zhao, yuhang.zhao@cs.wisc.edu, Department of Computer Science, University of Wisconsin-Madison, Madison, Wisconsin, United State; Chun Yu, chunyu@mail.tsinghua.edu.cn, Department of Computer Science and Technology, Tsinghua University, Beijing, China; Yuanchun Shi, shiyc@mail.tsinghua.edu.cn, Department of Computer Science and Technology, Tsinghua University, Beijing, China.

2474-9567/2022/9-ART126
https://doi.org/10.1145/3550290

and an observation study (N=12). We proposed a series of design guidelines targeting high precision, high flexibility, and low workload in interaction, and iteratively designed and developed a reading prototype application. Our implementation supports dynamic text structure and is supplemented by both linear and non-linear layered text navigation. The evaluation results (N=8) show that compared to existing methods, our prototype improves the clarity of text understanding and fluency of revisiting with reduced workload.

CCS Concepts: • **Human-centered computing → Empirical studies in accessibility**; *Accessibility systems and tools*; Empirical studies in interaction design.

Additional Key Words and Phrases: Accessibility, Visual impairment, Revisiting in reading, Design guidelines, Touchscreen, Interaction design, Navigation methods, Reading tools

ACM Reference Format:
Zhichun Li, Jiang Yu, Xiaochen Liu, Yuhang Zhao, Chun Yu, and Yuanchun Shi. 2022. Enhancing Revisitation in Touchscreen Reading for Visually Impaired People with Semantic Navigation Design. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 6, 3, Article 126 (September 2022), 22 pages. https://doi.org/10.1145/3550290

1 INTRODUCTION

There are at least 285 million people worldwide experiencing a visual impairment or blindness, and many of them are children and teenagers [46]. Yet, currently many educational resources are limited to visual access [15, 43], posing barriers to people with visual impairments since they need haptic (braille books or displays) or audio information (listening to audiobooks) to access those resources. The development of screen readers and Text-To-Speech technology on mobile phones [17] with abundant online resources opens up new reading opportunities for the visually impaired (VI) community [20, 48]. However, current reading methods on touchscreens still can't satisfy their advanced needs in academic reading, especially *revisiting*, the process of returning to previously-visited regions during reading.

Revisiting is a regional, quick, and non-linear text manipulation method typically achieved by selective jumps, the path of which vary from person to person. It occurs when readers wants to consolidate their memory, settle confusions, and seek topical information about the text [34, 41, 42]. As revisiting helps gain more thorough understanding of the text, it's a frequent and vital technique in intensive reading [3, 6, 50].

Past studies show that sighted and VI readers alike rely on spatial clues and the corresponding structural control of the text to navigate in revisiting tasks [19, 28, 29]. Yet spatial information is lost in touchscreen based reading for VI people due to the linearized audio content [2]. Flexible and effortless structural navigation is also extremely limited in current eyes-free touchscreen reading apps as only fixed-duration and linear jumps are supported in audio streams. Past researches have demonstrated effective ways to enhance navigation in different contexts - by supplementing spatial information like providing physicality of paper books through screen overlays [11, 13, 14, 24], or by supporting structural navigation by leveraging the conceptual structures of the websites [36, 49]. However, there lacks a thorough investigation of how to enhance navigation for revisiting tasks in touchscreen-based VI reading.

To address this, we investigated and augmented VI users' revisiting experience during touchscreen reading. We identified VI's behaviors, strategies, and challenges during revisiting by conducting an online survey (N=37) and an observation study (N=12). Based on the responses and interviews, we summarized four key factors (reading focus, text granularity, interaction methods, and confirmation) in revisiting as well as the needs and preference for each factor.

Building upon these findings, we proposed design guidelines for eyes-free touchscreen reading and revisiting. Our guidelines leveraged dynamic, interaction-based text reorganization and prioritized accuracy, flexibility, and low workload during the text revisiting process. Following the guidelines, we collaboratively designed a touchscreen reading tool with VI readers and implemented a prototype application, Semantic Reader for Visually

Impaired (SRVI). To investigate how our application augments the revisiting process, we conducted an evaluation with eight visually impaired participants (N=8), which demonstrated that SRVI significantly improved VI users' revisiting experience and enabled them to revisit with lower workload and an improved fluency. This paper can both benefit the VI touchscreen reading experience and inspire eyes-free mobile interaction design.

This paper's main contributions are three-fold:

- Through an online survey and an observational study, we identified four key factors of VI's revisiting behavior and investigated the challenges of VI revisiting on touchscreens.
- We proposed design guidelines that prioritized dynamic text segmentation and semantic navigation design for audio reading and revisiting on touch screen devices.
- We built SRVI, a Chinese-based mobile reader for VI. Our evaluation shows that VI readers were able to revisit more efficiently, flexibly, and comfortably on SRVI compared to existing reading apps.

2 RELATED WORK

2.1 Access to Reading for the VI Community

The mainstream reading tools currently used by the visually impaired can be divided in two categories: reading through touching (braille books or braille displays) and reading through listening (audio streams). Compared to conventional braille devices, audio-based VI reading on touchscreens is more portable and resourceful due to the ease of sharing audio contents [5, 13]. Thanks to the growing popularity of mobile devices and support services like screen reader and text to speech technologies, audio-based VI reading has been made more accessible. As a result, the VI community has been increasingly using smartphones and touchscreens with screen readers for daily reading [8, 17].

The change of sensory channel results in barriers during revisiting. Past empirical studies have identified two main differences that contribute significantly to challenges in audio reading. During the transition from braille to audio, the two-dimensional nature of the text is linearized into one dimension. Thus many important macro- and micro- spatial structures originally provided by haptic clues in braille devices are lost [7, 26, 39]. The linearized, single-threaded audio and limited interaction also force the readers to attentively and continuously listen to the narration instead of reading at their own pace as they could on a braille device [23]. This has contributed to the high cognitive load as well as worse text comprehension when reading through audios on touchscreens[38, 44].

To investigate how well current tools support VI touchscreen audio-based readings, we surveyed some common reading applications available for VI including Chinese-supported apps like WeChat Reading, Douding, VoiceDreamReader, etc..

We found that these applications support VI touchscreen audio-based reading in two ways: either with a screen reader (text-based) or through audiobooks generated in advance (timeline-based). For the text-based method, the app imports the text which is synthesized by the screen reader to speech and read aloud. The book or document is treated and manipulated the same way as other textual modules on the screen and therefore lacks a text-based and flexible control. For the timeline-based method, users only have control over the timeline which is structurally and semantically decoupled from the actual reading progress in the text.

2.2 Revisiting in reading

As Alexander [3] defined, revisiting is the strategy of returning to previously visited regions, which is a frequent activity in some documents. Marshall and Bly [6] observed that "looking back to re-read" is one of the four important types of lightweight navigation when people read paper materials, they will return for context information, like a name or other missed details. Revisiting is generally assumed to be needed due to comprehension failures in a casual or initial reading [35]. As an action that occurs randomly in time and with varying locational targets, revisiting requires a flexible movement among different scales of text granuity [30, 40].

A past study shows that similar to sighted readers, VI readers rely on spatial information and the corresponding spatial control of the text to help with their navigation in revisitation [19, 28, 29]. For instance, in braille documents text is organized in straight lines with indentations to identify distinct blocks of text, and braille letters are printed alongside them for labels and captions [12, 22]. Studies conclude that the layout of segmented texts of different units across space allows comprehensive and adaptive control over the temporal aspect of reading [16, 26, 45], allowing people to read non-linearly. Such navigation based on touchscreens, however, is poorly supported in current touchscreen eyes-free reading apps (See 2.1) or past studies targeting revisiting.

Past studies on either sighted or visually impaired revisiting were conducted on other tools which leverage very different interaction mediums (e.g., keyboards and mouses) [3, 36, 49, 50] compared to touchscreen mobile devices. Yet there lacks an investigation of the eyes-free revisiting experience on touchscreen devices which is crucial to enhance VI mobile reading and revisiting. We thus fill this gap by conducting an observational study. We confirmed the significance of manipulation flexibility and spatial information in navigation for revisiting, and extended this implication to touchscreens for the VI community.

2.3 Supporting and Augmenting Revisiting on Touch Screens

Generating bookmarks is one of the common methods to assist revisiting yet it fails to support random revisiting needs. Manually adding bookmarks is impractical as readers don't know their revisiting targets during initial reading. Some studies proposed automatic approaches to monitor users' reading behavior and thereby to predict revisiting needs. For example, bookmarks are generated automatically by tracking how long users stay on the current page [3] or supervising the user's electrodermal activity [31]. However, these approaches rely on the system's predictions and thus are inflexible and unfit for random revisiting needs. Furthermore, managing the list of bookmarks incurs manipulation cost and influences the revisiting experience[1]. Therefore, a flexible in-text navigation method is still needed so that VI readers can navigate quickly and freely to their random revisiting positions.

Some studies have demonstrated that providing structural or semantic information enhances navigation and thus revisiting. Machulla [26] suggested that a clear connection between logical document structures and navigation should be taken into account when designing non-visual reading interfaces. Yang and Rohani [36, 49] also proved that screen readers could navigate on websites more efficiently by leveraging the semantic structure of website pages. We investigate, leverage and further extend this established importance of semantic navigation in our study of touchscreen eyes-free reading and revisiting. Based on a layered text structure, our proposed design guidelines leverage semantic navigation to achieve more flexible and efficient control of the text.

The benefits of incorporating haptic and audio feedback to aid navigation in reading have also been demonstrated. Advanced Auditory Menus [47] designed four types of auditory scroll bars different in tone. Asakawa [4] emphasizes single words and phrases through tactile vibrations. Others [11, 13, 14, 23, 24] provided physicality of paper books through screen overlays, thereby manipulating haptic feedback for richer spatial information. Inspired by these works we thus investigated how to design and incorporate unobtrusive feedback in our study, and applied them to enrich the semantic and spatial information provided by the system.

These past solutions are instructive to our work, yet they are unfit for our purpose since they fail to fulfill one or more of the following requirements: based on touchscreens, VI-friendly, non-intrusive (without external accessories), and enhanced revisitation. We thus aim to fill this gap by proposing a VI touchscreen based reading and revisiting scheme.

3 AN OBSERVATIONAL STUDY TO UNDERSTAND VI REVISITING ON TOUCHSCREENS

Since there lacks a thorough investigation on VI's reading and revisiting experiences on touchscreens, we conducted an observational study to understand their current situation.

3.1 Surveying Tools and Texts

To ensure the observational study would be carried out in more realistic conditions, we conducted an online survey, which VI people can access through screen readers, to investigate background information of VI's reading and revisiting.

3.1.1 Survey Structures. The survey questionnaire was structured as follows:

(1) **Common reading apps used by BVI readers:** After researching the top rated reading apps and discussing with some BVI readers, we summarized nine most commonly used reading apps by the BVI community. They are Douding, Voice Dream Reader, WeChat Reading, Youshu App, Dedao App, Himalaya, Dragonfly FM, Dopoda, and Knowledge Planet. Participants were asked to choose 1-3 apps that they commonly used or name an alternative app if not in the list, and to answer the following question (2) and (3) for each app.
(2) **Revisiting frequency in apps:** After introducing revisitation, participants were asked about their revisit frequency in their selected apps using Five-point Likert scale, anchored at "I never revisit" (1), "I sometimes revisit" (3), and “I always revisit" (5).
(3) **Types of text in apps:** We adopted Maslova's classification of text types [27] in terms of their communicative intentions. The text types include narration (telling the story), description (descriptive writing involving imagery and specific details), exposition (writing that tells the reader or explains the subject), and argumentation (state the author's point of view and try to persuade the reader). Participants were asked to choose the text types they commonly read on each app as well as the ones they would read in a hypothesized reading app that supports fluent revisiting.
(4) **Demographics:** We collected standard social demographic information including gender, age, vision, the highest education level currently in progress or obtained.

3.1.2 Participants. To reduce differences in understanding and individual capabilities due to education background or age, we recruited visually impaired adults aged 20 to 30 with an educational level above college degree in all experiments. Participants could access the surveys with their screen readers of their choices. We retrieved 37 surveys (23 males and 14 females, age M=24.59, SD=3.77) as 3 disqualified surveys were deleted due to misunderstanding or mistakes in filling in the survey. Of all participants, 13.5% have associate degrees, 81.08% have bachelor degrees, and 5.41% have master degrees.

3.1.3 Results. The survey results are as below.

The apps used by BVI users vary across individuals while WeChat Reading and Himalaya Reading are the most popular ones. 64.84% of the participants use WeChat Reading and 67.57% use Himalaya Reading. 86.49% of the participants use at least one of WeChat Reading and Himalaya Reading. Other seven apps rated between 2.7% to 13.51%. Due to lack of users for other apps, we focused on analysing the using condition of these two apps as the primary BVI mobile reading apps.

Currently VI readers revisit, but not frequently, in mobile reading. The frequency of revisiting in WeChat Reading (M=3.25, SD=0.52) and Himalaya Reading (M=2.88, SD=0.66) indicates that revisiting happens under current mobile reading conditions. Though the frequency might also be influenced by devices and personal reading habits, this indicates that there are behaviors of revisiting using existing apps. That is, people do have a need to revisit when reading on touchscreens.

Supporting revisiting motivates VI users to read more diverse text types, especially informational ones. Table 1 and 2 shows the text types that users of WeChat Reading (N=24) and Himalaya (N=25) would like to read on different apps. An optimized revisiting functionality raised willingness to read expositions by 34% and 32%, raised narrative and argumentative texts by 12% and 25%, 32% and 28%, and reduced the willingness to read descriptive texts by 21% and 4% in WeChat Reading and Himalaya respectively. This implies that the lack of support in revisiting in these two apps hinders users from reading non-descriptive texts, especially informational ones like expositions.

Table 1. Proportion of text types that WeChat reading users tend to read in different apps.

Applications	Narration	Description	Exposition	Argumentation
WeChat Reading	70.83%	87.50%	33.33%	33.33%
Hypothesized App with Fluent Revisiting	83.33%	66.67%	62.50%	66.67%

Table 2. Proportion of text types that Himalaya users tend to read in different apps.

Applications	Narration	Description	Exposition	Argumentation
Himalaya	44.00%	76.00%	36.00%	20.00%
Hypothesized App with Fluent Revisiting	76.00%	72.00%	64.00%	68.00%

3.2 Investigation Method

We conducted an observational study to understand VI readers' revisiting behavior, challenges, and needs in both timeline-based and text-based audio reading.

3.2.1 Apparatus. Considering functionalities and users' preferences, WeChat reading was selected for the observational study. Though with a similar usage rate among the survey participants, WeChat reading supports both text-based and timeline-based reading while Himalaya Reading mostly relies on audio based podcasts and audio books. Furthermore, WeChat Reading users have higher revisiting frequency which aids our observation.

Participants were asked to use the WeChat Reading app on their phones to simulate reading and revisiting scenarios in daily life (See Figure 2). Based on their daily reading habits, they can either (1) use the screen reader to read text in the app (text-based), or (2) listen to the built-in audiobooks (timeline-based). Interaction functions provided by method (1) include play, turning pages, and adjusting text segmentation granularity to letter/word by a rotor. Turning pages is enabled by Wechat Reading and all the other functions are provided by the screen reader. Functions provided by method (2) include play/pause, dragging the progress bar, and jumping forward/backward in the audio stream by 15s, all the functions are provided by WeChat Reading app.

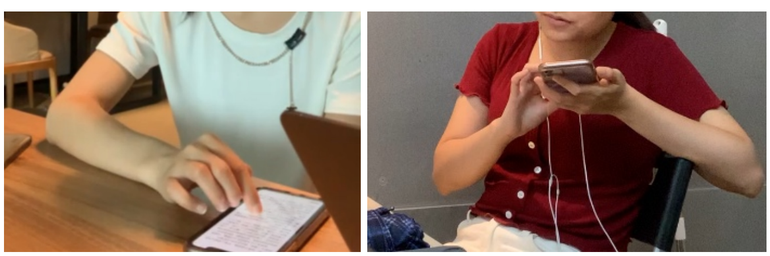

Fig. 2. VI participants are using WeChat Reading.

3.2.2 Participants. Through convenience sampling with the same criteria as the survey, we recruited 12 visually impaired participants (5 females and 7 males) aged from 20 to 28 (M=22.66, SD=2.90). Five participants were completely blind, five had light perception, one had low vision, and one had secondary visual disability. All of them were experienced in using WeChat Reading. Their personal information is detailed in Table 7. The entire experiment lasted around 120 minuets, and each participant was compensated 40 dollars.

3.2.3 Text. Since the survey results in Section 3.1.3 indicated that the increase of willingness to read a certain text type is the most significant in expository texts if fluent revisiting is supported, we selected a text-only segment from a book named *Criminal Law Handout* written by Law professor, Xiang Luo. We chose a chapter of length 2601 Chinese characters that explains the origin of the penalty. As there lacks a standard measurement of text readability [21] in Chinese, we conducted a pilot study with 2 VI readers to ensure proper readability and difficulty of text.

3.2.4 Task. Participants were asked to imagine that they were planning to take an extra law course, and needed to read this article to learn some basic knowledge ahead of time. They should aim to learn and understand the text as much as possible to perform well in class.

Revisit occurs randomly in location and in time at varying distances. Thus to thoroughly observe and understand revisiting in different scenarios, we asked four text understanding questions, two of which were asked during reading to simulate the need for instant revisiting; and the other two were asked shortly after initial reading to simulate revisiting at a longer distance or that seeks topic information. The questions mainly asked about the reasoning logic of the text, and were unlikely to be answered upon initial reading. Participants were encouraged to revisit but would be asked to stop the task when a 3-minute limit was reached as the time would allow us to thoroughly observe how participants move within the text to revisit.

3.2.5 Procedure. Participants first answered demographic questions (age, gender, vision, education level, and reading habits). Then they were asked to read the selected text using WeChat Reading in a specified scenario. A following personal interview was conducted asking about how they feel about their revisiting experience on WeChat Reading, and what interaction functionalities they used for revisiting. To encourage the participants to share more about their experience, focus groups of three were conducted in which participants were led to share their past revisiting experience on Braille displays or computers, and their ideas of an ideal revisiting experience on touchscreens. This study was approved by the university's Institutional Review Board.

3.3 Findings

We thoroughly observed the VI participants revisiting behavior through the tasks. When a need for revisiting occurred, the participants started the revisiting process with a revisiting target - a position in the text that they hoped to reach. The VI participants typically started by moving their reading focus in larger scales around the text based on their impression upon initial reading. The jumping patterns depended largely on the text manipulation functionalities provided by the device/app and the participants' personal habits. For example on text-based revisiting, some jumped by words and others preferred to continuously read; on timeline-based revisiting, some jumped by 15 second segments while others dragged the progress bar. Participants then zoomed in to smaller regions either by performing smaller jumps or listening through the text content to search for the exact revisiting target. During this whole process the participants' text manipulation path, though following the above described general pattern, was random, impacted by previous understanding, and differed a lot in individuals. This implies the need for interaction flexibility in supporting revisiting for different people.

Through our observation we summarized four key factors of revisiting, the improvement of which might help enhance the revisiting experience. To start with, while adapting paperback books to texts or audio on touch screens, these apps restructure the content with different **text granularities** - pages, audio time lengths, etc.

Relying on this bottom-level structuring of the text, the readers' **reading focus** is then tracked at the scale in correspondence with the text granularity. In moving the reading focus in the text, **interaction methods** transform readers' gesture inputs to the according text manipulation results. Due to their impaired visions, **awareness** of both the reading progress and the user input is crucial to readers. This provides feedback which is useful in helping readers locate themselves correctly and thus guiding them to revisit more efficiently. These four factors thus support each other and are strongly interrelated.

After video-recorded the whole process, we qualitatively coded participants' response to identify a series of VI readers' challenges and needs in revisiting on mobile phones through our observations, interviews and focus group discussion. Three researchers first discussed the categories and themes of data, and worked separately for the open coding process, and then ran a semantic analysis to get a consensus after a second round discussion. Based on these responses we list and explain the identified challenges and needs.

3.3.1 Inflexible and semantically disruptive text segmentation. Five participants reported dissatisfaction towards the text segmentation methods which are inflexible and semantically disruptive. Text segmentation in text-based method is often constrained by the screen's graphical layouts (e.g., pages and lines). During the interview, four participants (P1, P2, P5, P10) complained about unnatural interruptions as the audio continuously played through pages. Though they recognized that such interruption gave spatial clues of *"they are on a new page"*, it interrupted reading and their semantic understanding of the text. For the timeline-based method, the text was restructured into 15-seconds audio segments which was also semantically disruptive.

3.3.2 The inappropriate text segmentation limited the revisiting path. Four participants reported that the logic of segmentation limited their revisiting path. Specifically, user navigated linearly by page, line (only TianTan screen reader on Android), word or letter in text-based style. If the user wanted to locate a word at the middle or the end of a page, he/she had to navigate using small rotors like lines and words continuously, which wasted plenty of time to hear the non-target text. P8 said that *"This process is very troublesome and I hope it can be optimized, such as enabling jumps"*. In the focus group discussion, all four groups expressed their hopes to add navigation methods of different scales, as they commented: *"I think semantic content units of different scales will be useful, rather than only being able to turn single pages."*

3.3.3 Inaccessible interaction design adds to workload. All participants accumulated varying degrees of dissatisfaction towards the apps' inaccessible interaction design through their past experience and during the experiments. For instance, the interface for audiobooks provided by these apps were not designed specifically for VI readers thus still rely heavily on visual cues for text manipulation. While using the audiobooks, VI readers can either jump by 15-seconds audio segments, or by dragging a progress bar in an unsegmented text. However, the crucial information such as the position of the bar, time length moved, and the entire time length of the audio is not available. For VI readers, it's extremely hard or cognitively burdensome to establish the correspondence between physical dragged length on the progress bar and the text length moved without the visual clues. P7 commented during the interview that *"I tried to use the progress bar to search and locate keywords, but I did not know where to drag it"*.

3.3.4 Easily lost reading focus due to graphical user interface (GUI) based focus tracking. All participants indicated the importance of knowing their location throughout the reading process, yet reported that they frequently lost their reading focus during the revisiting process. This was caused by GUI-based reading focus tracking.

Existing apps such as WeChat Reading does not discriminate text content with other control modules on the screen and only one interaction focus is allowed on the same user interface at one time. If users touch other controls mistakenly during reading, the reading focus in the text content is lost consequently. Readers then have to spend a large amount of time relocating the focus, which was reported to be *"exhausting"* (e.g. P4).

3.3.5 Navigation throughout the text with some granularities is constrained. Three participants found they were blocked when they arrived the start or the end of the screen page by using the word rotor, since currently the screen reader only tracks the reading focus within one screen page. Users had to filp a page to continue reading and their reading focus was reset to the start of the new page, resulting in a lose of their perceptual reading focus. VI readers thus were unable to navigate throughout the text flexibly with the preferred text granularity and unconstrained by the screen-based segmentation. P2 reported this *"disturbs my comprehension and slows down the revisiting process".*

3.3.6 Lack of locational information of reading focus. During focus group discussion, nearly all participants in focus group discussion indicated that their best revisiting experience was obtained on braille books because the reading focus, which is where the readers are touching, could move freely (e.g. Group 1). When reading on braille books, VI readers' reading progress, or the location of the reading focus, can be easily known by perceiving the current approximate page number by the thickness of the book in the three-dimensional space with the position on the page in the two-dimensional space touched by the finger (e.g. Group 4). While braille books are rarely used for daily reading due to limited resources and the difficulty to carry (e.g. Group 2), nearly all participants indicated that they hope to extend such ease of obtaining information of the reading focus onto touch screen reading. This is currently lacking in mobile reading due to the lack of reading progress feedback.

Despite three groups (Group 1, 2 and 4) mentioning the search function in revisiting, it is time consuming and only applies to revisiting exact words or phrases, thus limiting. As P8 also pointed out, *"Searching involves many steps, and I need to sift through a list of results. It's very time consuming. I prefer to navigate to the text location quickly."*

4 DESIGN GUIDELINES

We propose a series of design guidelines to be taken into account in future VI mobile reading design. The guidelines are concluded from: (1) the results of our observation study (Section 3.3); (2) an investigation into theoretical and empirical research on revisiting (Section 2.2); and (3) study of eyes-free design interaction in audio-based reading (Section 2.3). The guidelines aim to optimize VI users' revisiting experience in terms of high accuracy, high flexibility, and low workload.

4.1 DG1: Semantic and dynamic text structure

In VI mobile reading, the two-dimensional books are linearized into a singular audio stream. Semantic and structural information, including words, paragraphs, pages, etc., is thus lost [7, 26, 39]. This poses difficulties in both understanding and interaction. It is thus paramount to add semantic information back into the linear audio stream. Text structures of different granularities should also adapt dynamically to the users' interaction intentions.

4.1.1 Semantically aware and dynamic text segmentation at varying granularities. We discussed in Section 3.3.1 that current VI reading Apps segments the text according to the device's graphical user interfaces(GUI), resulting in lack of semantic information and abrupt semantic interruptions during reading. Thus we argue that instead of segmenting the text based on GUI, which is inaccessible to the VI community, the text should be segmented in a semantically meaningful way and into different granularities such as words, sentences, paragraphs, pages, and chapters. Furthermore, to support smooth transitioning between different granularities, the text should be dynamically restructured during the VI users' reading process.

4.1.2 Mapping segmented texts of different granularities to enable precise tracking of the user's reading position. Section 3.3.5 discussed that users were thus unable to navigate throughout the entire text flexibly with the same or changing text granularities. A precise and flexible tracking based on the dynamically and semantically segmented text content instead of the GUI thus should be supported. This requires establishing correspondences between each pair of units in differently processed texts that are semantically segmented by words, sentences, paragraphs, etc. Users should be able to read with any granularity and to switch between granularities smoothly without unexpected focus shift. They should not be required to navigate with the screen page granularity when reading by words on different pages; reader' reading focus should also remain unchanged when switching to a different text granularity.

4.2 DG2: Semantic Navigation and VI-friendly Interaction

The semantic and dynamic text segmentation in DG1 augments the originally linear audio stream, opening up new opportunities to fulfill users' text manipulation needs. Correspondingly, VI or eyes-free text manipulation should adapt to different levels of text granularities flexibly while also prioritizing the needs and preferences of the VI community in the interaction experience.

4.2.1 Separating the text from other controls to stably track the reading focus. In Section 3.3.4 we observed that users would lost their reading focus when clicking on other interaction modules. We suggest to separate the reading focus from other modules on the GUI to avoid unintentional lose or shift of the focus. The reading process should be treated as a standalone process, which should not be influenced by other text-irrelevant interactions such as opening the setting menu.

4.2.2 Layered text manipulation. In correspondence to the supported text segmentations at different granularities in DG1, interaction methods should be layered, concise, and intuitive. Users should be able to leverage distinct interaction methods for navigation with differently segmented texts. For example, users should be able to easily navigate to the last/next sentence, last/next word, etc.

4.2.3 Non-linear text navigation. Section 3.3.2 also showed that currently users could only navigate linearly. Yet As the process of revisiting is comprised of regional, irregular, and selective jumps to the target position (Section 3.3). It is thus important to support non-linear navigation throughout the text and at all levels of text segmentation. Jumping between sections is a typical example of non-linear navigation on a large scale. VI users should be able to jump non-linearly at different scales to help with improving the revisiting efficiency.

4.2.4 VI user-friendly interaction methods with minimal cost. Section 3.3.3 indicates the VI readers' needs for accessible and VI-centered interaction design. Visually impaired users are more likely to complete gestures with relatively low requirements of position or speed. They usually avoid using gestures that are demanding in accuracy, and favor interaction with tangible hints like those at the edge or the corner [18]. These VI interaction behaviors and preferences need to be taken into account when designing low-requirement, accessible, and comfortable interaction methods for VI mobile reading. The interaction design should ensure that the minimum possible cost would incur when users mistouch. This helps easy retraction of unintended module activation.

4.3 DG3: Unobtrusive feedback for input confirmation and contextual information

Since the probability of mistouch exists, uncertainties in the the user input - text manipulation correspondence occur due to lack of input confirmation. Such uncertainties intensify for VI users who are interacting quickly with the text and don't have visual feedback. Also, due to the speciality of the audio streams, users can only obtain singular and linear information at one time. This leads to the lack of comprehensive understanding of users' reading status. Therefore feedback for input confirmation and contextual information is crucial in helping users to understand and control their reading status. Choosing the appropriate form of feedback should also be taken into account.

4.3.1 Feedback for input confirmation. Unlike tangible books on which text manipulations like page flips can be confirmed and easy to amend if an error occurs (by flipping the page once again), users' gesture inputs on touch screens only result in changes in the audio streams. It is thus hard to detect mistouch or to confirm if the gesture input suits the user's intentions. Thus to establish a strong and clear correspondence to every pair of gesture input and the resulting textual operations to avoid confusion and to timely amend the mistouch, feedback is helpful and essential. Feedback can be provided in forms of audio or haptic feedback to provide input confirmation to help users understand their reading status better. For example, users should be prompted when they flip a page so users can adjust their interactions accordingly if they did it mistakenly.

4.3.2 Feedback for contextual information. Section 3.3.6 discussed that tangible books conveniently provide contextual information through the thickness of the part of the book read, position in the current page, etc. Yet without the visual channel and due to the lack of tangible information of screen displays, contextual information is lost in VI mobile reading which makes it hard for users to establish a cognitive map [9, 25]. This information should thus be supplemented through providing active and passive feedback. For instance, some subtle passive feedback can be provided locally and contextually at the end or the beginning of a sentence, a paragraph, a page, etc. to remind users of their progress as they read. Active feedback can be provided upon the user's request to tell more detailed and comprehensive information, such as the user's current position in the entire chapter or book.

4.3.3 Unobtrusive feedback. All feedback methods that supplement the reading process should be secondary compared to the primary text content audio flow. This implies that any form of either audio or haptic feedback should not disturb the reading process. The strengths and limitations of the feedback modality should be taken into account when choosing feedback for different scenarios. For instance, auditory feedback, which include textual information, can convey more details yet would disturb the ongoing reading flow. On the other hand sound and haptic feedback have less conflict with the text audio yet they convey limited information.

5 AN IMPLEMENTATION WITH SRVI

We embodied the guidelines (Section 4) for designing an eyes-free reading app named SRVI (See features in Table 5 and 6). SRVI structures the text dynamically based on semantic text segmentation of varying granularities and incorporates linear and non-linear navigation at different text granularity accordingly. SRVI also leverages user-friendly and unobtrusive audio and vibration feedback to support tracking of the reading focus which is defined as the exact point in the text that the user is reading.

The app was developed to investigate the effectiveness of our proposed design guidelines for eyes-free revisiting. To achieve an intuitive gesture-functionality correspondence, we designed the gestures based on past studies that proposed guidelines for establishing such correspondence in eyes-free interaction [33, 37], Voice User Interface [10, 32], and gestures for VI [18]. In deciding on the optimal layout and the most effective feedback in realizing the functionalities, we iteratively developed 3 versions of the app and conducted pilot studies with fourteen visually impaired people in total to refine our designs.

Dynamic and adaptive text segmentation. SRVI supports users' interaction with the text at different granularities by dynamically adapting the text segmentation based on user gesture inputs. The text is segmentated based on semantic granularities instead of text length (DG1-1) so users can make seamless transfer to the intended granularity based on their semantic understanding. We preprocessed the text into three granularities (word group based, sentence based, and page based) and established the structural correspondence between each pair to enable smooth transfer (See Fig 3). Thus in SRVI these three segmentations are available for text manipulation. The segmentation, as well as the reading focus (DG1-2), adapts dynamically to user input commands. The switch between different granularities is simply performed by a command of a specific granularity, e.g. a next sentence/next page input. Meanwhile, the complexity of the input command gesture input coincides conveniently

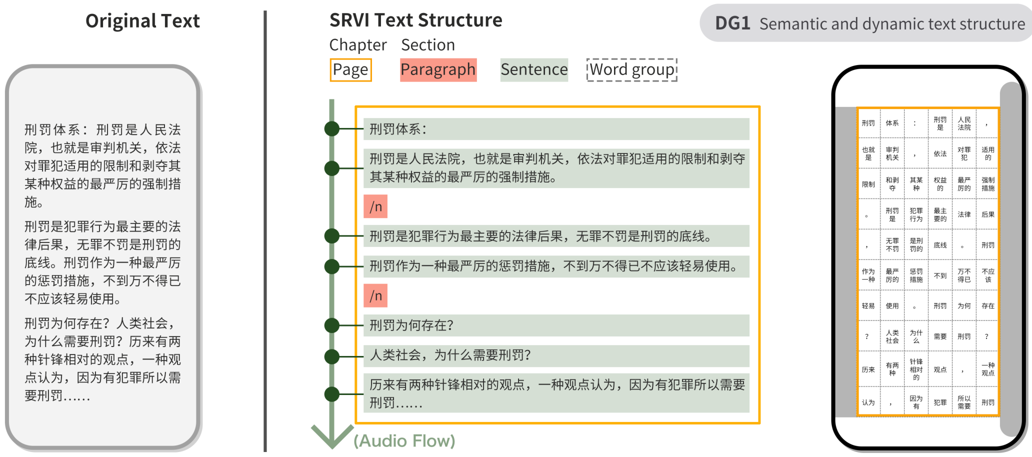

Fig. 3. The text structure in SRVI based on a semantic segmentation in Chinese, guided by DG1.

with the granularity of the text segmentation (one finger gestures for sentences and two finger gestures for pages).

Standalone text stream with progress following read-head. SRVI separates the text stream and its read-head from all other input commands received on the screen (DG2-1). Any non-text input command (reading status report, line number report, etc) can't affect the text read-head position to avoid accidental loss of read-head. The read-head accurately follows the user's reading position with optimal precision. The read-head linearly flows with the playing audio and jumps in correspondence with the user's active text manipulation gesture input. In realizing such a design, SRVI employs dynamic and adaptive text segmentation so that SRVI's text read-head automatically adapts to the exact word on the screen, the start of a sentence, or the start of a page. This gives the user flexibility in manipulating the text at their desired granularity.

Bimodal text manipulation towards integrated linear and non-linear navigation. Both reading and revisiting behaviors involve sequentially reading through the text with varying granularities (DG2-2) and quick, selective jumps to a specific text region (DG2-3). In achieving this we incorporated both linear and non-linear navigation at different segmentation granularities (See Fig 4). Linear navigation is achieved through basic functionalities like last/next sentence or page. Non-linear jumps are achieved by a progress bar and a grid-based touch-read mode to support jumps at varying scales. The progress bar, which was designed to be a sliding bar of the length of the screen, allows users to jump to a page within the current chapter. Compared to the current audio-track based progress bar, SRVI's progress bar allow spatial navigation which simulates turning to a specific page in paper books.

To provide smaller scale word jumps, which helps users understand specific words when reading a complex text, we leveraged a bimodal design. Other than listening to the text under a listen-read mode, touch-read mode is activated when auto-playing stops. The touch-read mode is automatically disabled when auto-playing resumes to avoid mistouch. A grid based design on the screen allows the user to perform small jumps within the current page. The displayed grid self-adapts to phones of different sizes and on our test phones the display is a 6 by 10 grid. Considering SRVI users' preference for limiting gesture interactions to edges and corners(DG2-4), we reserved a fixed-length non-text region around the grid for gesture interaction. Straying are highly likely in touch-reading mode as VI readers may unconsciously and unintentionally stray to a neighboring line without visual information. A past study remedy this by automatically adapting word content to users' touching path

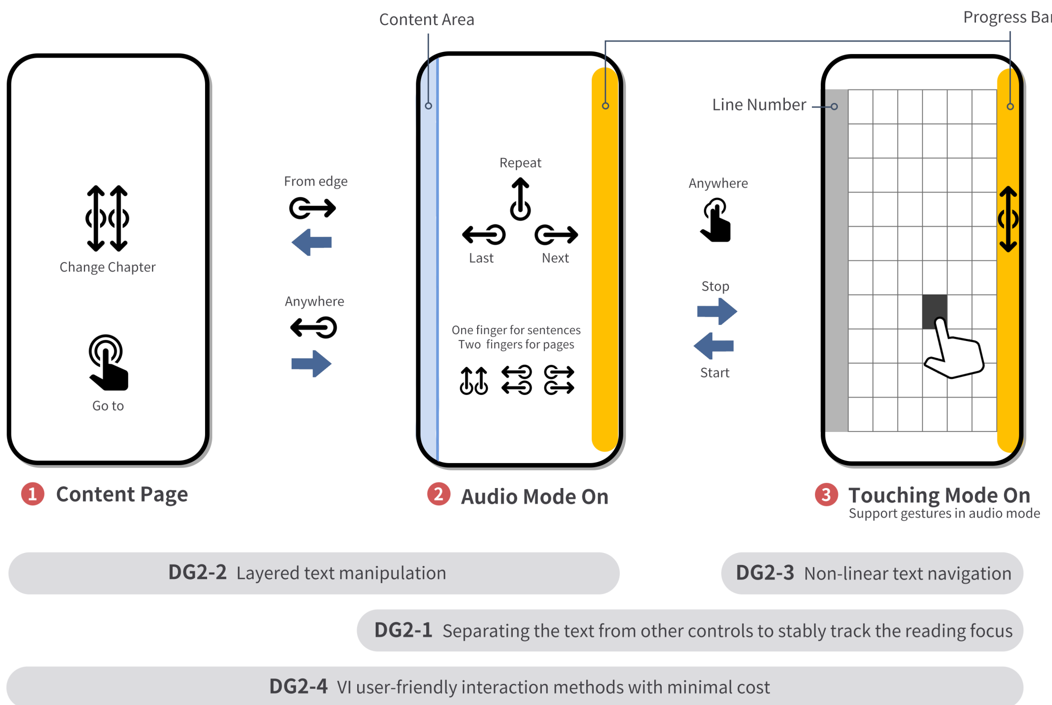

Fig. 4. Interaction methods designed based on DG2. The Audio mode page (2) is the main interface, where users can listen continuously. When the reading stops, touching mode (2) is active, which allows users to touch-read word groups in grids. Both (2) and (3) support navigation by gesture commands and the adaptive progress bar.

on the screen [11]. We employed a different approach and used an algorithm to detect straying behavior. Once straying is detected strong and long vibration feedback is provided as an warning for unintentional line crossing. Readers could then be guided to return to the right line. Our method preserve the spatial layout in a page which readers may rely on for navigation.

Input confirmation and progress awareness through audio and vibration feedback. Users need a definite and clear association between their active gesture inputs to the text being played such that unexpected text jumps that don't conform to inputs lead to confusion and less confident control of the text. SRVI thus provides audio or vibration feedback as gesture input confirmation in ways that don't interfere with the text audio (DG3-1, 3, see Fig 5). In facilitating readers to perform selective jumps with our bimodal system, users are actively informed by audio reports of their current selection when performing jumps. Specifically, the system reads out which page or chapter the user will be jumping to when sliding on the progress bar or the menu; the left side of the touch-read grid reads out the line number when touched to inform users their current position on the screen. Sound feedback is provided when flipping a page (page flip sound) and entering into the selected chapter from the menu (locking sounds), etc. Vibration feedback is provided as pseudo haptic boundary outlines and line-crossing warnings when touch reading word groups on the screen. Short vibrations are provided whenever a vertical or horizontal line in the grid is crossed so that users are aware of their current position on the screen; longer vibrations are provided when the system detects an unintentional line crossing characterized by a slow and gradual movement downwards and towards the right.

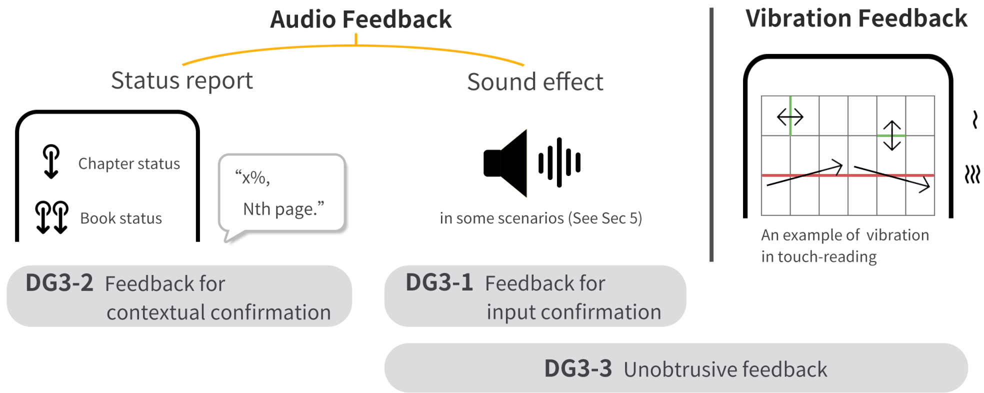

Fig. 5. An illustration of the audio and vibration feedback design in SRVI, which is guided by DG3.

Audio feedback is also provided upon users' request to report their current locations within the book or the chapter (DG3-2). This is meant to provide a global and comprehensive view of the structure of the entire text which is fundamental to locating the revisiting spot quickly.

SRVI was implemented as an Chinese based android app developed in Android Studio with the audio content synthesized by Iflytek TTS engine, a popular Chinese TTS engine.

6 EVALUATION OF SRVI ON REVISITING TASKS

In order to verify the applicability of our design guideline and to further explore the influence of the design of reading tools on revisiting, we evaluated users' revisiting performance on SRVI in comparison with WeChat Reading. WeChat Reading was included as a reference for participants' daily revisiting experience. The experiment provided potential insight for a better solution in the future.

6.1 Method

We conducted the study in a controlled lab setting and used a 2 (tools) × 2 (articles) within-subjects design. In total, participants will complete the task 2 times using SRVI and WeChat Reading. The articles and the tools are matched according to the Latin Square design and the order of the conditions is counterbalanced.

6.1.1 Participants and Apparatus. We recruited 8 visually impaired participants (4 females and 4 males) aged from 20 to 25 (M=23.00, SD=1.51) by doing convenience sampling. Five participants were completely blind, one had light perception, and one had level one visual disability. They all have rich experience in using WeChat Reading. The participants of this study do not overlap with the previous observational study. The entire experiment lasted around 100 minutes, and each participant was compensated 35 dollars.

All participants used WeChat Reading on an iPhone 11 with VoiceOver. The functionalities available remain the same as in the previous study. SRVI runs on the Huawei Mate 40 Pro with Android 10 without a screen reader supplement. The speaking rate of Huawei mobile phones was set to match that of WeChat Reading.

We continued to use the *Criminal Law Handout* from Section 3.2.3. Chapter 3 and Chapter 33 are selected with 2,090 and 2,627 words.The two chapters express the importance of the criminal law constraining the state and the system of criminal law.

6.1.2 Procedure. Participants were first taught and given enough time to familiarize themselves with the operations. Participants took an average of 11.59 minutes (SD = 2.88) to familiarize themselves with SRVI. We tested

Fig. 6. VI participants using two types of apps, from left to right: SRVI, WeChat Reading

Table 3. Quantitative results.

Metrics	SRVI	WeChat Reading
Completion rate	90.63%	87.50%
Time (s)	51.88 (SD=17.08)	88.71 (SD=26.02)

and confirmed their capability to use SRVI by a series of basic interaction tasks. Then they were asked to imagine studying on their own in preparation for a pretest for a law course. They needed to put effort into reading and understanding to get good grades in the exam. All eight participants completed the three revisitation tasks two times counterbalanced with two texts using SRVI and WeChat Reading. Three questions would be asked during (two questions) the and after (one questions) minutes of the reading process to simulate different types of revisit in real life. Participants would be stopped once they revisited over three minutes. After the tasks were finished, they used Likert-scale, open interview, and NASA-TLX to conduct a comprehensive evaluation of the two apps. This study was approved by the university's Institutional Review Board.

6.1.3 Data Collection and Analysis. We analyzed whether SRVI outperformed WeChat Reading in terms of time and completion rate on revisitation tasks. Since users could always find the answer without time limits, we defined a task as incomplete when the time limit was surpassed. We rejected the cases where participants answered the question from memory without revisiting. We conducted a paired t-test to analyze the difference in time. We used ANOVA to evaluate different aspects of revisiting experiences of the two apps by Likert-scale, and used NASA-TLX to measure participants' workloads of the two apps. The behavioral-based open interviews asked participants to provide their overall feelings, the difficulties encountered, suggestions on revisions, and their overall feelings about using those apps to complete tasks. After the tasks, a semi-structured interview was conducted. We collected video and audio records and observation notes as qualitative data.

6.2 Results and Findings

6.2.1 Quantitative results. The completion rate for SRVI (90.63%) was similar to WeChat Reading (87.50%), while participants spent significantly less time ($t = 2.668$, $p<0.05$) on the SRVI (M=51.88, SD=17.08) than the WeChat Reading (M=88.71, SD=26.02). The difference in time indicates that SRVI brings about significant improvement on revisiting efficiency.

P5, P7 and P8 each had one incomplete task when they used WeChat Reading, while P5 had two incomplete tasks when using SRVI. Since revisit requires cues established from the pre-revisiting memory of content and its location in the text, when participants have no impression of target text at all, the help provided by different tools is limited. As P5 said: *"I have no impression of this content at all, so I don't know where to look for it."* Participants acknowledged that tools wouldn't be able to help much in this case. *"I may have been distracted when I was listening to that position. I don't remember mentioning it."* (P7)

6.2.2 Qualitative results. While doing revisiting tasks, the **workload** was significantly lower ($t=-2.459$, $p<0.05$) on SRVI (M=28.53, SD=13.17) than on WeChat Reading (M=40.86, SD=19.89). Looking at each indicator of NASA-TLX in detail, compared to WeChat Reading, SRVI was rated lower on Frustration and Mental, slightly improved on Physical and Temporal, remained similar on Performance and Effort. This indicates that participants did not need to spend much energy on manipulating the text when using SRVI, and the revisiting process was more in line with their expectations. They also believed that SRVI reduced the workload of revisiting and improved the efficiency. Since most participants could complete the task through revisiting, they had a relatively close evaluation of their achievements and efforts.

The **subjective ratings** in table 4 shows that SRVI outperformed WeChat Reading on all four factors and the entire experience of revisiting (proposed in Section 3.3). The only indicator where WeChat Reading was rated better than SRVI was learnability, but the difference was not significant. The reason was that all participants were proficient in WeChat Reading while just start using SRVI, and SRVI added far more functions and gestures. Participants commented during the interview that SRVI was the best mobile reading app they have used and they were willing to use this app for daily reading.

Table 4. Subjective Score in evaluation experiment (1 = Strongly Disagree, 5 = Strongly Agree).

Subjective Metrics	**Statements**	**SRVI**	**WeChat**	**F and p value**
Learnability	The app is very easy to learn.	4.00 (0.93)	4.50 (0.93)	1.167 (>0.05)
Facilitate revisit	It improves my revisiting experience.	4.75 (1.71)	4.50 (0.93)	18.180 (<0.05)
Reading Focus	Accurate, hard-to-lost reading focus.	4.13 (1.13)	2.38 (1.41)	8.763 (<0.05)
Text Granularity	Appropriate segmentation for movement.	3.88 (1.55)	2.50 (1.07)	7.568 (<0.05)
Interactive Experiences	Interaction experiences of revisiting is good.	4.50 (0.93)	2.25 (1.49)	7.000 (<0.05)
Confirmation	Easy to confirm my reading location.	4.75 (0.46)	3.00 (1.31)	11.949 (<0.05)

6.2.3 Qualitative Experience. **Semantic segmentation is the most appreciated text structuring method.** Participants pointed out the similar problems of WeChat Reading to Section 3.3, that is, it segments text by layouts (See Section 5 and 6.1.1) which cuts off complete sentences during navigation, resulting in semantic disruption and confusion. Moving by semantically segmented units such as sentences in SRVI better their understanding of the content, and the cooperation of features made VI users manipulating text more accurately. P6 gave an example in the interview, commented that *"it's easy to repeat a sentence on SRVI and I don't have to listen to redundant content"*; while on WeChat Reading users had to relocate the sentence by re-listening almost the entire page.

In addition, word group segmentation based on semantics was also thought to be helpful. P1 appraised the touch reading of semantic word segmentation, he explained: *"The words in each grid are meaningful. When I turn to a new page, I can touch a few random boxes and recall the outline of the page based on the key words I heard."*

SRVI provides abundant confirmation information which greatly reduced users workload on locating their current reading focus and knowing their reading status. WeChat reading shortly pauses the playing audio stream when a page is turned, which provides confirmation to some extent but was also reported to be disruptive to the reading process and would interrupt the coherence of the user's understanding of the article. In comparison, participants generally considered SRVI's sound feedbacks more informative and unintrusive. *"It is more like I am reading a real book."* (P2) We observed that when P4 and P7 used SRVI, they would broadcast the current page number before revisiting, which enabled them to quickly return to the original reading position after revisiting. However, when using WeChat Reading, participants had to go back and forth to find the previous reading focus, which is time consuming and is extremely frustrating. A participant commented that *"broadcasting position is a very good function."* (P7)

Personal choices on navigation methods vary. Similar to the findings of the observational study, users exhibited personally distinct reading behaviors in the evaluation. That is, users chose to use different functionalities in varying orders and at different times. This was most evident in the use of the touch-reading mode and the progress bar. Touch reading is favored by some people because they like the freedom of reading wherever they touch (See Fig 7). They also commented that this method would play a more prominent role in academic material where they need to know everything literally (P3). Others explained their infrequent use of touch-reading

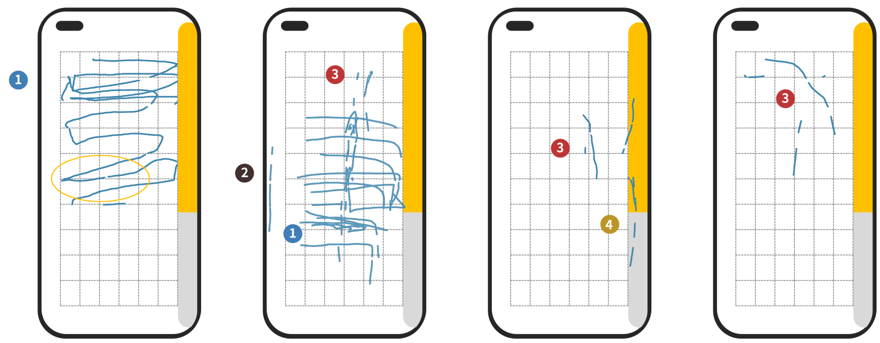

Fig. 7. The tracks of four participants' fingers on a certain page when revisiting. With the video record of the experiment, we found four types of strategies in touch-mode: (1) general sequential touch-reading, (2) read the number of line to locate. (3) non-linear movements to find keywords, continue to read the current page if find (case B) and leave if not (case C and D), and (4) use progress bar to jump. Case A also shows a directional correction after being prompted with vibration feedback.

feature by unfamiliarity or unawareness of the feature during the reading or revisiting process. The same thing also happened for the progress bar. Many users were impressed by the location information and flexibility of movement it provides, while others rarely used it because they are not used to such long jumps.

7 DISCUSSION

7.1 Contributing to the Research on VI Reading and Revisiting

This paper aims to enhance the revisiting experience in VI mobile reading on touch screens. The evaluation on our implementation SRVI shows that compared to existing common mobile reading methods, users completed revisiting tasks more efficiently and with lower workload on SRVI. This implies that our implementation based on the proposed design guidelines in Section 4 is effective. Our work enriches the research on touchscreen reading for VI people, especially the process of revisiting. The design guidelines we proposed and implemented (SRVI) supplemented previously lacking flexible and eyes-free revisiting in mobile reading.

Based on the observation study, we restructured the text semantically on different scales to reduce the lost of two-dimensional textual clues in linearized audio stream. The semantic, multi-level, and dynamic text segmentation supports flexible text navigation and was proven in our evaluation to enable personalized text manipulation for different VI readers. The rich variety of interaction commands and feedback that we implemented in SRVI was preferred by the participants compared to existing apps like WeChat Reading. The evaluation also showed that VI participants adapted some of their braille book reading strategies to reading in SRVI, such as touch-reading word groups non-linearly (See Fig 7), which was desired by VI readers yet improbable in existing touch-screen reading tools (Section 3.3).

We do recognize that SRVI is a stage towards fully supporting revisiting on touch-screens based on a subset of the design guidelines that we proposed and thus can be further optimized. With our investigation, we hope to inspire other implementations based on our design guidelines as well as extensions of the guidelines themselves.

7.2 Adding Non-Text Dimensions Into VI Mobile Reading

In this paper, we chose a text-only segment in implementation SRVI as we focused on adding two-dimensional semantic textual clues from books back into linearized audio streams, which is fundamental to reading. A book, however, contains more than spatial information - tables and figures are oftentimes provided for illustrative purposes, text colors and fonts are used to emphasize and highlight important information, etc. Though not relevant to the textual information, these factors aid text understanding and thus guide revisiting. To transform and adapt the multi-dimensional visual information to haptic and audio on flat touch screens, further study on VI readers' preferences in receiving these information is needed. We will explore how to supplement the text with other information such as tables, figures, and text colors in VI or eyes-free mobile reading in the future.

7.3 Generalizing the Design Guidelines to Other Languages

Our implementation, SRVI, based on the proposed design guidelines was in Chinese. Because the guidelines are based on semantic text segmentation certain implementation details of SRVI are highly relevant to language such as the grouping of characters into word groups in Chinese. Yet though the detailed segmentation methods differ we argue that since semantics exist in all languages, the design guidelines can be adapted to other languages to support VI revisiting on touch-screens.

7.4 Compatibility with Screen Readers

Most existing mobile reading tools for the VI community leverages built-in screen readers. SRVI, which supports semantic text navigation, was implemented without screen readers due to their GUI-based content segmentation and navigation. We recognize that screen readers are useful when navigating through GUI-based contents and layouts. The screen readers can thus be used in non-text contents in further touch-screen reading tools, such as book shelves and settings, etc. The control would then be transferred over to a semantic-based navigation tool, such as SRVI, when the VI user starts reading to support flexible and intuitive text manipulation.

7.5 Limitation and future work

We recognise some of our limitations and propose to continuously refine our work in the future.

To reduce the impact of the cognitive load of maneuvering unfamiliar gestures in the evaluation, participants were given time to learn and familiarize themselves with the interaction gestures in the evaluation. It's demonstrated that they were able to learn and adopt the gestures within a short time. Yet during the short reading and revisiting time, participants were unable to establish new mobile reading habits or proficiently maneuver all the interaction functionalities provided by SRVI. Many participants reported that they believe SRVI would be more helpful once they are more familiar with it. To avoid the influence of fatigue and reduced attention towards a text, we conducted short-term evaluation on short segments of text on each participant. We chose a law-related text as the reading material in our evaluation study, supplementing the evaluation with different text types might help us more thoroughly and robustly evaluate SRVI. Due to the above two considerations, we plan to continue our evaluation in a long-term field study so that VI readers can read more diversely and interact with the App more proficiently.

We chose highly educated VI participants in our study because as they were more likely to read academic texts in which revisiting might occur frequently. We also plan to recruit participants with more diverse education background to investigate whether their revisiting and text navigation behaviors differ.

8 CONCLUSION

In this paper, we addressed the problem of revisiting reading on touchscreens for VI people. We identified the revisiting behaviors and challenges of VI readers through an observational study based on which we proposed a set of design guidelines. We then implemented a prototype by iteratively designing and developing SRVI, a reading app with VI-friendly, semantic navigation methods. We evaluated the performance of SRVI in a lab setting. Though users' preference on specific features differ, the qualitative and quantitative results showed that SRVI helped VI readers revisit more efficiently, flexibly and frequently, and were highly appreciated on the subjective experience.

Our work focused on supplementing linearized audio streams with two-dimensional semantic clues to help text segmentation. Other than optimizing SRVI based on the guidelines, we aim to explore adapting more dimensions (text colors, table, etc.) of visual-based reading to visually impaired mobile reading. This would help augment the information available to visually impaired readers and make more resources accessible.

ACKNOWLEDGMENTS

This work is supported by National Key R&D Program of China under Grant No.2020AAA0105200, Beijing Academy of Artificial Intelligence (BAAI), and the Natural Science Foundation of China under Grant No.62132010. Our work is also supported by Beijing Key Lab of Networked Multimedia, the Institute for Guo Qiang, Tsinghua University, Institute for Artificial Intelligence, Tsinghua University (THUAI).

REFERENCES

[1] David Abrams, Ron Baecker, and Mark Chignell. 1998. Information archiving with bookmarks: personal web space construction and organization. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 41–48.

[2] Faisal Ahmed, Andrii Soviak, Yevgen Borodin, and IV Ramakrishnan. 2013. Non-visual skimming on touch-screen devices. In *Proceedings of the 2013 international conference on Intelligent user interfaces*. 435–444.

[3] Jason Alexander, Andy Cockburn, Stephen Fitchett, Carl Gutwin, and Saul Greenberg. 2009. Revisiting read wear: analysis, design, and evaluation of a footprints scrollbar. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 1665–1674.

[4] Chieko Asakawa, Hironobu Takagi, Shuichi Ino, and Tohru Ifukube. 2002. Auditory and tactile interfaces for representing the visual effects on the web. In *Proceedings of the fifth international ACM conference on Assistive technologies*. 65–72.

[5] Lyra C Bertulfo, Lhercy Anne A Cotoner, Jhaztin M Namit, Alvin Claude V Pacheco, Ma Corazon G Fernando, and James C Felizardo. 2017. Gabay tinig: a 3d interactive audiobook with voice recognition for visually-impaired and blind preschool students using mobile technologies. In *Proceedings of the 3rd International Conference on Communication and Information Processing*. 99–103.

[6] Sara Bly and Catherine C Marshall. 2005. Turning the page on navigation. In *Proceedings of the 5th ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL'05)*. IEEE, 225–234.

[7] Yevgen Borodin, Jeffrey P Bigham, Glenn Dausch, and IV Ramakrishnan. 2010. More than meets the eye: a survey of screen-reader browsing strategies. In *Proceedings of the 2010 International Cross Disciplinary Conference on Web Accessibility (W4A)*. 1–10.

[8] Scott Carter and Laurent Denoue. 2009. SeeReader: An (Almost) Eyes-Free Mobile Rich Document Viewer. *arXiv preprint arXiv:0909.2185* (2009).

[9] Marvin M Chun and Yuhong Jiang. 1998. Contextual cueing: Implicit learning and memory of visual context guides spatial attention. *Cognitive psychology* 36, 1 (1998), 28–71.

[10] Michael H Cohen, Michael Harris Cohen, and Giangola. 2004. *Voice user interface design.* Addison-Wesley Professional.

[11] Gurjot S Dhillon, Yasmine N El-Glaly, William H Holbach, Tonya L Smith-Jackson, and Francis Quek. 2012. Use of participatory design to enhance accessibility of slate-type devices. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 56. SAGE Publications Sage CA: Los Angeles, CA, 1952–1956.

[12] Nicoletta Di Blas, Paolo Paolini, Marco Speroni, et al. 2004. Usable accessibility" to the Web for blind users. In *Proceedings of 8th ERCIM Workshop: User Interfaces for All, Vienna*. Citeseer.

[13] Yasmine N El-Glaly and Francis Quek. 2014. Digital reading support for the blind by multimodal interaction. In *Proceedings of the 16th International Conference on Multimodal Interaction*. 439–446.

[14] Yasmine N El-Glaly and Francis Quek. 2016. Read what you touch with intelligent audio system for non-visual interaction. *ACM Transactions on Interactive Intelligent Systems (TiiS)* 6, 3 (2016), 1–27.

[15] Danna Ethan and Charles E Basch. 2008. Promoting healthy vision in students: progress and challenges in policy, programs, and research. *Journal of School Health* 78, 8 (2008), 411–416.

[16] John DE Gabrieli and Elizabeth S Norton. 2012. Reading abilities: importance of visual-spatial attention. *Current biology* 22, 9 (2012), R298–R299.

[17] William Grussenmeyer and Eelke Folmer. 2017. Accessible touchscreen technology for people with visual impairments: a survey. *ACM Transactions on Accessible Computing (TACCESS)* 9, 2 (2017), 1–31.

[18] Shaun K Kane, Jacob i19O Wobbrock, and Richard E Ladner. 2011. Usable gestures for blind people: understanding preference and performance. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 413–422.

[19] Keefer.R, Liu.Y, and Bourbakis.N. 2013. The Development and Evaluation of an Eyes-Free Interaction Model for Mobile Reading Devices. *IEEE TRANSACTIONS ON HUMAN-MACHINE SYSTEMS* 43, 1 (2013), 76–91.
[20] Taslim Arefin Khan, Dongwook Yoon, and Joanna McGrenere. 2020. Designing an eyes-reduced document skimming app for situational impairments. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–14.
[21] J Peter Kincaid, Robert P Fishburne Jr, Richard L Rogers, and Brad S Chissom. 1975. *Derivation of new readability formulas (automated readability index, fog count and flesch reading ease formula) for navy enlisted personnel.* Technical Report. Naval Technical Training Command Millington TN Research Branch.
[22] Georgios Kouroupetroglou and Dimitrios Tsonos. 2008. Multimodal accessibility of documents. *Advances in Human-Computer Interaction* (2008), 451–470.
[23] Kyung-Ryong Lee, Beom Kim, Junyoung Kim, Hwajung Hong, and Young-Woo Park. 2021. ADIO: An Interactive Artifact Physically Representing the Intangible Digital Audiobook Listening Experience in Everyday Living Spaces. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–12.
[24] Jingyi Li, Son Kim, Joshua A Miele, Maneesh Agrawala, and Sean Follmer. 2019. Editing spatial layouts through tactile templates for people with visual impairments. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–11.
[25] Liang-Yi Li, Gwo-Dong Chen, and Sheng-Jie Yang. 2013. Construction of cognitive maps to improve e-book reading and navigation. *Computers & Education* 60, 1 (2013), 32–39.
[26] Tonja Machulla, Mauro Avila, Pawel Wozniak, Dillon Montag, and Albrecht Schmidt. 2018. Skim-reading strategies in sighted and visually-impaired individuals: a comparative study. In *Proceedings of the 11th Pervasive Technologies Related to Assistive Environments Conference*. 170–177.
[27] T. Maslova. 2015. MODES OF DISCOURSE IN THE ENGLISH LANGUAGE CLASSROOM. (2015).
[28] Millar.S. 1997. Reading by Touch. *Routledge* (1997).
[29] Meredith Ringel Morris, AJ Bernheim Brush, and Brian R Meyers. 2007. Reading revisited: Evaluating the usability of digital display surfaces for active reading tasks. In *Second Annual IEEE International Workshop on Horizontal Interactive Human-Computer Systems (TABLETOP'07)*. IEEE, 79–86.
[30] Kenton O'hara and Abigail Sellen. 1997. A comparison of reading paper and on-line documents. In *Proceedings of the ACM SIGCHI Conference on Human factors in computing systems*. 335–342.
[31] Matthew KXJ Pan, Joanna McGrenere, Elizabeth A Croft, and Karon E MacLean. 2014. Exploring the role of haptic feedback in enabling implicit hci-based bookmarking. *IEEE transactions on haptics* 7, 1 (2014), 24–36.
[32] Peter Parente. 2006. Clique: a conversant, task-based audio display for GUI applications. *ACM SIGACCESS Accessibility and Computing* 84 (2006), 34–37.
[33] Lara Schibelsky G Piccolo, Ewerton M De Menezes, and Bruno De Campos Buccolo. 2011. Developing an accessible interaction model for touch screen mobile devices: preliminary results. In *Proceedings of the 10th Brazilian Symposium on Human Factors in Computing Systems and the 5th Latin American Conference on Human-Computer Interaction*. 222–226.
[34] Anthony K Pugh. 1978. *Silent reading: An introduction to its study and teaching.* Heinemann Educational Publishers.
[35] Rayner and Keith. 1997. Understanding Eye Movements in Reading. *Scientific Studies of Reading* 1, 4 (1997), 317–339.
[36] Romisa Rohani Ghahari, Mexhid Ferati, Tao Yang, and Davide Bolchini. 2012. Back navigation shortcuts for screen reader users. In *Proceedings of the 14th international ACM SIGACCESS conference on Computers and accessibility*. 1–8.
[37] Uday Sagale, Ganesh Bhutkar, Mahadev Karad, and Nikhil Jathar. 2018. An eye-free Android application for visually impaired users. In *Ergonomics in Caring for People*. Springer, 291–297.
[38] Bastian Schildbach and Enrico Rukzio. 2010. Investigating selection and reading performance on a mobile phone while walking. In *Proceedings of the 12th international conference on Human computer interaction with mobile devices and services*. 93–102.
[39] Venkatesh Sivaraman, Dongwook Yoon, and Piotr Mitros. 2016. Simplified audio production in asynchronous voice-based discussions. In *Proceedings of the 2016 CHI conference on human factors in computing systems*. 1045–1054.
[40] Robert Spence. 1999. A framework for navigation. *International Journal of Human-Computer Studies* 51, 5 (1999), 919–945.
[41] Craig S Tashman and W Keith Edwards. 2011. Active reading and its discontents: the situations, problems and ideas of readers. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2927–2936.
[42] Alexander Thayer, Charlotte P Lee, Linda H Hwang, Heidi Sales, Pausali Sen, and Ninad Dalal. 2011. The imposition and superimposition of digital reading technology: the academic potential of e-readers. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2917–2926.
[43] Carolina Cumani Toledo, AP Paiva, Gustavo Bittencourt Camilo, MR Maior, IC Leite, and Maximiliano Ribeiro Guerra. 2010. Early detection of visual impairment and its relation to academic performance. *Rev Assoc Med Bras* 56, 4 (2010), 415–9.
[44] Kristin Vadas, Nirmal Patel, Kent Lyons, Thad Starner, and Julie Jacko. 2006. Reading on-the-go: a comparison of audio and hand-held displays. In *Proceedings of the 8th conference on Human-computer interaction with mobile devices and services*. 219–226.
[45] Erik Wästlund, Torsten Norlander, and Trevor Archer. 2008. The effect of page layout on mental workload: A dual-task experiment. *Computers in human behavior* 24, 3 (2008), 1229–1245.

[46] WHO. 2021. World report on vision 2021. http://www.who.int/mediacentre/factsheets/fs282/en/ Document.
[47] Pavani Yalla and Bruce N Walker. 2008. Advanced auditory menus: Design and evaluation of auditory scroll bars. In *Proceedings of the 10th international ACM SIGACCESS conference on Computers and accessibility.* 105–112.
[48] Shunguo Yan and PG Ramachandran. 2019. The current status of accessibility in mobile apps. *ACM Transactions on Accessible Computing (TACCESS)* 12, 1 (2019), 1–31.
[49] Tao Yang, Prathik Gadde, Robert Morse, and Davide Bolchini. 2013. Bypassing lists: accelerating screen-reader fact-finding with guided tours. In *Proceedings of the 15th International ACM SIGACCESS Conference on Computers and Accessibility.* 1–8.
[50] Yuhang Zhao and Yongqiang Qin. 2014. QOOK: enhancing information revisitation for active reading with a paper book. In *Proceedings of the 8th International Conference on Tangible, Embedded and Embodied Interaction.* 125–132.

Table 5. Interaction design in the reading interface.

Function	**Finger quantity**	**Starting position requirement**	**Action**	**Tactile feedback**
Open content	1	left edge	right fling	N/A
Start reading (enter listen-reading mode)	2	anywhere	single tap	N/A
Stop reading (enter touch-reading mode)	2	anywhere	single tap	N/A
Last/next sentence	1	anywhere	right/left fling	N/A
Last/next page	2	anywhere	right/left fling	N/A
Back to sentence start	1	anywhere	up fling	N/A
Back to page start	2	anywhere	up fling	N/A
Chapter status report	1	anywhere	down fling	N/A
Book status report	2	anywhere	down fling	N/A
Progress bar	1	right edge	press, scroll, release at right edge	when scrolled to a new page (crossing a gridline)
Cancel progress bar	1	right edge	left fling	N/A

Function	**Auditory**	
	Invoked under listen-reading mode	Invoked under touch-reading mode
Open content page	"open menu"	
Start reading (enter listen-reading mode)	N/A	plays the text until the end
Stop reading (enter touch-reading mode)	click sound effect	N/A
Last/next sentence	plays last/next sentence, stops when sentence finished	
Last/next page	plays last/next page with flip page sound effect	"nth page"
Back to sentence start	replays current sentence	page reset to where the start of the current sentence is at
Back to page start	replays current page	N/A
Chapter status report	"x%, nth page" (as in this chapter)	
Book status report	"x%, nth page" (as in entire book)	
Progress bar	"to nth page"	starts reading the selected page
Cancel progress bar	"back to nth page"	

Table 6. Interaction design in the content page.

Function	Finger quantity	Starting position requirement	Action	Auditory	Tactile feedback
Close content	1	anywhere	left fling	"close menu"	N/A
Change chapter	1	anywhere	press, scroll, release	"nth chapter" + title	when scrolled to a new chapter (crossing a gridline)
Go to chapter	1	anywhere	double tap	lock sound effect	N/A

Table 7. Demographic information of the 20 BVI participants from observation and evaluation study. All information was self-reported.

No.	Age	Gender	Visual Condition	Education	Observation	Observation Group	Evaluation
1	21	F	light perception	Bachelor	P1	G1	
2	21	M	light perception	Bachelor	P2		
3	21	M	weak light perception	Bachelor	P3		
4	21	M	blind	Bachelor	P4	G2	
5	20	F	low vision	Bachelor	P5		
6	21	M	blind	Bachelor	P6		
7	22	M	weak light perception	Bachelor	P7	G3	
8	28	M	light perception	Bachelor	P8		
9	22	M	blind	Bachelor	P9		
10	26	F	weak light perception	Master	P10	G4	
11	21	F	blind	Bachelor	P11		
12	28	F	blind	Bachelor	P12		
13	24	F	blind	Bachelor			P1
14	24	F	motion perception	Bachelor			P2
15	20	M	blind	Bachelor			P3
16	23	M	blind	Bachelor			P4
17	23	M	blind	Bachelor			P5
18	25	F	blind	Bachelor			P6
19	22	M	weak light perception	Bachelor			P7
20	23	F	light perception	Bachelor			P8